A Layered Approach to Constrain Signing Avatars

Paritosh Sharma

▶ **To cite this version:**

Paritosh Sharma. A Layered Approach to Constrain Signing Avatars. VISIGRAPP_DC 2023, Scitevents, Feb 2023, Lisbon, Portugal. hal-04143663

HAL Id: hal-04143663
https://hal.science/hal-04143663v1

Submitted on 27 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Layered Approach to Constrain Signing Avatars

Paritosh Sharma[1][a]
[1]*LISN, CNRS, Université Paris–Saclay, Orsay, France*
paritosh.sharma@lisn.upsaclay.fr

1 RESEARCH PROBLEM

In human communication, the sign languages are the main languages used by deaf people around the world. Synthesis of these sign languages is a promising method for deaf communication, allowing us to customize and create new sign language content and preserve the artist's anonymity.

Triggering pre-recorded animations from a gloss-based database is a common method for synthesizing sign language (Pezeshkpour et al., 1999). Here, a gloss represents the sign and is mapped to a clip of an avatar performing the sign. However, this technique requires a lot of time and manpower to create these pre-recorded clips. Therefore this method does not scale up well with large sign language utterances and cannot be applied to avatars in virtual worlds where maintaining large databases of animations is not possible.

These problems of scaling and data management have motivated research into the synthesis of sign language with procedural methods (GIBET et al., 2001). Here, a gloss is mapped to a sequence of motion constraints to be evaluated and synthesized on the avatar. Nonetheless, synthesizing realistic motion with such systems remains a difficult problem, and addressing the signer's prosody, expressivity, and identity by providing control over style is even more challenging.

Glosses have traditionally been used as a formalization of sign language utterances. Yet this imposes the problem of synchronisation and reusability of those signs, which vary with a change in context.

To solve this, the AZee model (Filhol et al., 2014) allows us to write parameterised signed forms for semantic functions. Given a description, it generates a timeline that specifies every aspect of the utterance that the avatar should produce, resolving the problems with timing, sign concurrency, and non-manual features synchronization. Additionally, interpolation information is contained in AZee's temporal specifications, which is crucial for synthesizing the utterance.

This has motivated research for data-driven synthesis from AZee (Filhol et al., 2017). A synthesized utterance can be depicted as a set of blocks on a muti-track timeline (Sharma and Filhol, 2022). These blocks can be generated using evaluated low-level posture constraints or pre-recorded animations. However, all low-level constraints were generalized as a set of Inverse Kinematics(IK) Problems to solve. For specific scenarios, relying on the IK and joint limits to constrain movement of the posture is not enough. Thus, we introduce a layer-based approach to solving constraints and show how it can be used for a complete data-driven sign language synthesis model.

[a] https://orcid.org/0000-0001-9938-008X

2 OUTLINE OF OBJECTIVES

The objectives of this work can be summarized as follows:

- Show problems with only IK-based constraining.
- Present a new layered-based approach to better synthesize AZee constraints.

3 STATE OF THE ART

Animation from AZee descriptions can be divided into two categories: pre-animated and bottom-up synthesis(building from minimal constraints). Pre-animated methods use explicit, often manually created, mappings of utterance description to motion data. (Filhol and Mcdonald, 2018) also uses template utterance descriptions and facilitates the generation of utterances with parameterized motion sequences. However, the diversity of these generated utterances is limited to the number of designed animations and database content. Moreover, the required manual labour hinders scalability.

The Bottom-Up synthesis creates motion from minimalist constraints rather than relying on pre-animated mappings. Despite recent research efforts, the naturalness of generated motion still falls significantly behind that of a pre-animated motion. However, it provides a broader coverage since it doesn't

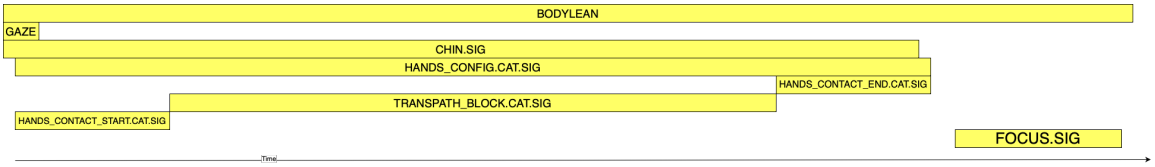

Figure 1: Arrangement of the utterance for expression in section 3 on the timeline

rely on pre-recorded motion data and hence, can produce all motion specified by the utterance description.

The AZee native level defines several basic types to constrain an armature posture. To understand it better, let's consider the following AZee expression from the corpus *40 brèves* (Challant and Filhol, 2022) (Filhol and Tannier, 2014) (LIMSI and LISN, 2022)

```
:about-point
    'pt
    ^Rssp
    'locsig
    :category
        'cat
        :pays
        'elt
        :Irak
```

The above AZee expression is a representation of an SL production assigning "Iraq" to a point on the right-hand side of the signing space, a way of creating a reference the signer can later point to refer to the country. Let's assume we have pre-recorded action for the rule *:Irak*. fig. 1 represents this utterance on a multi-track timeline. The blocks constrain the posture in the following ways:

- **HANDS_CONTACT_START.CAT.SIG**: Placement constraints to keep the fingertips in contact in the beginning.
- **HANDS_CONTACT_END.CAT.SIG**: Placement constraints to keep the fingertips in contact at the end.
- **TRANSPATH_BLOCK.CAT.SIG**: Transpath constraint specifying the movement of hands along an arc
- **HAND_CONFIG.CAT.SIG**: Constraints for finger configuration.
- **CHIN.SIG**: Constrain the chin up
- **GAZE**: Constrain the gaze to the right signing space
- **BODYLEAN**: Orient the back so it leans towards the right.
- **CHIN.SIG**: Constrain the chin up
- **FOCUS.SIG**: pre-recorded action for the rule *:Irak*

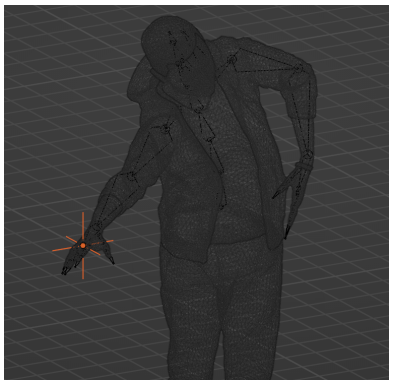

Figure 2: Hand IK chain pulling the spine and the shoulder

During block application, the IK chain to be chosen is based on the joint dependencies to evaluate the *placement* constraints. Though this system does allow for the extension of the IK chain through the spine, it invokes the solver with a new chain every time it applies a placement. This results in a slower evaluation of placements; moreover, the arms and torso have different purposes in the human body (fig. 2). The torso movement could have its own meaning irrespective of the movement of the arm. Lastly, whenever we switch the chain for the next constraint, we lose the IK information required by the other constraints, such as *transpaths*(movement specified along a path) or other placements.

To fix this, some systems define the spine, and arm IK separately (Baerlocher and Boulic, 2004) (Elliott et al., 2008). (McDonald et al., 2016) present the use of spine and shoulder extensions with an analytical hand IK model to address the timing of spine and shoulder movement separately. This allows for more natural bust movement. Avatar layers based on behaviour, skeleton and muscles were initially introduced by (Chadwick et al., 1989). This interests us since we aim to define our posture based on its behaviour w.r.t. linguistic constraints. Thus, in the following section, we propose a layer-based posture configuration to solve and apply the native AZee constraints.

Figure 3: Use of IK behaviour in HANDS_CONTACT_START.CAT.SIG block from fig. 1

4 METHODOLOGY

The goal of our method is to constrain our posture using layers and specify the relationship between the layers through the native AZee constraints. (Chadwick et al., 1989) define a layer as a conceptual simulation model which maps higher-level parametric input into lower-level outputs. Thus, using layers of the character related through the AZee constraints, we aim to add higher-level control over the low-level skeleton specification.

We define our posture using the following three behaviour layers:

4.1 IK Behaviour layer

The IK behaviour layer represents the motion specifications for the skeleton's arms and fingers. Having pre-defined chains in a layer also allows for better evaluation of *transpath* constraints. This layer encapsulates the chain movements for each arm bone and finger phalanges. When the IK is applied, a numerical IK solver generates the joint rotations for each bone in the chain. The mesh deformations are generated by weight painting (Mohr et al., 2003) based on a given skeletal state.

4.2 Forward Kinematics(FK) Behaviour layer

The FK behaviour layer constitutes the motion specifications for all FK bones. It is used for all the local or global bone rotation changes in the skeleton. The constraints *orient*, *rotate*, *trill*, *look* and *lookat* use the FK layer.

Figure 4: Use of FK behaviour in CHIN.SIG block from fig. 1

4.3 Morph Behaviour layer

Certain linguistic constraints are more suitable to be evaluated using pre-defined shape keys rather than as other constraints in the IK or FK behaviour layers. For example, the following AZee expression constrains the posture to close its little finger.

```
L_closed
azop
    'param$0
    'nodefault
    orient
        'dir
        !little
            ^param$0
            2
        'along
        oppvect
            dir
            !palm
                ^param$0
```

However, this method is slower since it generates vectors for constraining each of the joints in the finger. When used for each finger, a rule like fist_closed will generate 20 vectors for each joint. Thus, morphs make it more suitable to define a part of the sign language motion space. The following conditions can be considered when defining morphs,

- The morph action has a linguistic meaning(fist_closed, brow_raised, etc.)
- The morph action evaluates to local rotations on the skeleton or to some shape of the mesh independent of the skeleton layer.

Thus, the above expression can be rewritten as,

Figure 5: Evaluated morph for $little\_closed(w) = 0.5$

```
L_closed
azop
    'param$0
    'nodefault
    morph
        'little_closed
    ^param$0
        1.0
```

A set of morphs is pre-defined on the morph layer and then applied to the posture with the specified weight during block evaluation.

4.4 Ordering the constraints

Once the Score is generated, for each block, the constraints have to be sorted based on their dependencies which can be represented as a dependency graph. The topological sorting of this graph gives us a set of sorted constraints. These sorted constraints are then used to create the dependency graph of the blocks, which determines the order of block evaluation. fig. 6 shows us the blocks for fig. 1 with their first constraint and edges representing dependencies.

4.5 Updating the layers

Since each layer affects the low-level skeleton specification, the other layers have to be updated with the skeleton as well during the constraint application.

5 IMPLEMENTATION

5.1 Pre-animated Dataset and Data Preparation

Based on our defined behaviour layers, we created simple animations for rules such as *:Irak*, *:cuisine*, etcetera. To use this action dataset effectively, we map the varying behaviour layer control node to the parameters from the AZee expression. fig. 7 shows the rule *:Irak* while the body leans towards right signing space.

5.2 Blender add-on

We implement our animator as an add-on in Blender(v3.4) (Community, 2018). fig. 8 shows the Blender interface configured with the AZee animator addon. Its main components include:

(a) Properties
Modify inverse kinematics (IK) settings and animation layers.

(b) Viewport
Shows the 3D scene with the avatar.

(c) Non-linear Editor
To place all the baked blocks from the utterance.

(d) Action Editor
The third etc allows us to modify and visualize the generated actions as well as the pre-recorded animations.

(e) AZee editor
An editor to evaluate AZee expressions. It also includes settings for armature configuration, toggling constraints, managing body sites and defining global signing space and camera position.

We use the AutoRigPro (Artell, 2023) add-on to implement the posture layers since it has pre-defined IK and FK switching mechanisms important for updating our layers.

6 EXPECTED OUTCOME

Our implementation is still under development. However, the current implementation allows us to visualise the AZee block timeline and generate simple utterances using both pre-recorded animations and minimal constraints. Synthesis for the utterance in fig. 1 can be seen at the following link.

https://github.com/Paritosh97/phd/raw/master/grapp-2023/outcome.mp4

We see that the animator can synthesise AZee expressions using both bottom-up and pre-animated techniques.

7 STAGE OF THE RESEARCH

The theme of my PhD is the development of a synthesis system which can synthesize AZee descriptions of

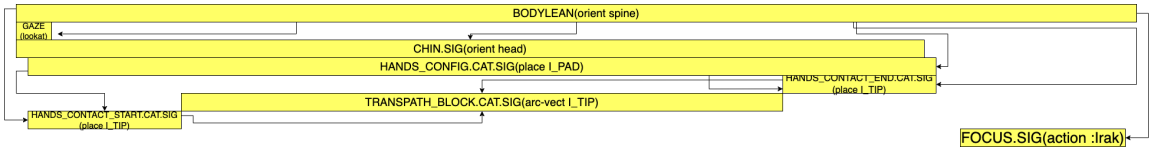

Figure 6: Block DAG for expression in section 3

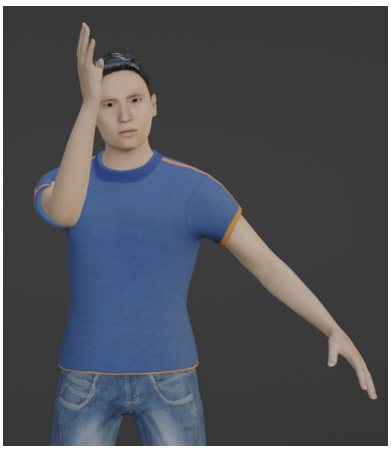

Figure 7: *:Irak* with body leaning towards the right signing space

a sign language discourse through a descending order of granularity. As explained earlier, we use pre-animated and bottom-up constraint synthesis for utterance generation. Thus, we can summarize the future goals of the PhD as follows.

- Improve usage of the pre-animated actions by parameterizing the motion data for the relevant specification given in the AZee expression.
- Increasing the quality of our bottom-up synthesis by increasing its naturalness using noise functions and better management of the f-curves using Bézier handles (Bechmann and Elkouhen, 2001).
- Integrating the above two techniques since the blocks generated using the bottom-up synthesis would look more robotic than those that used a pre-animated action. Here, applying the motion manifold from the pre-animated action data to solve the posture constraints can be a path to consider for a more seamless utterance generation (Holden et al., 2017).
- Testing and debugging our blender implementation for more complex utterances.

ACKNOWLEDGEMENTS

This work has been funded by the Bpifrance investment “Structuring Projects for Competitiveness” (PSPC), as part of the Serveur Gestuel project. Special thanks to my PhD director, Dr Michael Filhol, for providing guidance and feedback throughout this research.

REFERENCES

Artell (2023). Auto-rig pro.

Baerlocher, P. and Boulic, R. (2004). An inverse kinematic architecture enforcing an arbitrary number of strict priority levels. *The Visual Computer*, 20:402–417.

Bechmann, D. and Elkouhen, M. (2001). Animating with the “multidimensional deformation tool”. In Magnenat-Thalmann, N. and Thalmann, D., editors, *Computer Animation and Simulation 2001*, pages 29–35, Vienna. Springer Vienna.

Chadwick, J. E., Haumann, D. R., and Parent, R. E. (1989). Layered construction for deformable animated characters. *ACM Siggraph Computer Graphics*, 23(3):243–252.

Challant, C. and Filhol, M. (2022). A First Corpus of AZee Discourse Expressions. In *Language Resources and Evaluation Conference*, Proceedings of the 13th Language Resources and Evaluation Conference, Marseille, France.

Community, B. O. (2018). *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam.

Elliott, R., Glauert, J. R., Kennaway, J., Marshall, I., and Safar, E. (2008). Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Universal access in the information society*, 6(4):375–391.

Filhol, M., Hadjadj, M., and Choisier, A. (2014). Non-manual features: the right to indifference. In *International Conference on Language Resources and Evaluation*, Reykjavik, Iceland.

Filhol, M. and Mcdonald, J. (2018). Extending the AZee-Paula shortcuts to enable natural proform synthesis. In *Workshop on the Representation and Processing of Sign Languages*, Miyazaki, Japan.

Filhol, M., Mcdonald, J., and Wolfe, R. (2017). Synthesizing Sign Language by connecting linguistically structured descriptions to a multi-track animation system. In Margherita Antona, C. S., editor, *11th International Conference on Universal Access in Human-Computer Interaction (UAHCI 2017) held as Part of HCI International 2017*, volume 10278 of *Universal Access in Human–Computer Interaction. Designing Novel Interactions*, Vancouver, Canada. Springer.

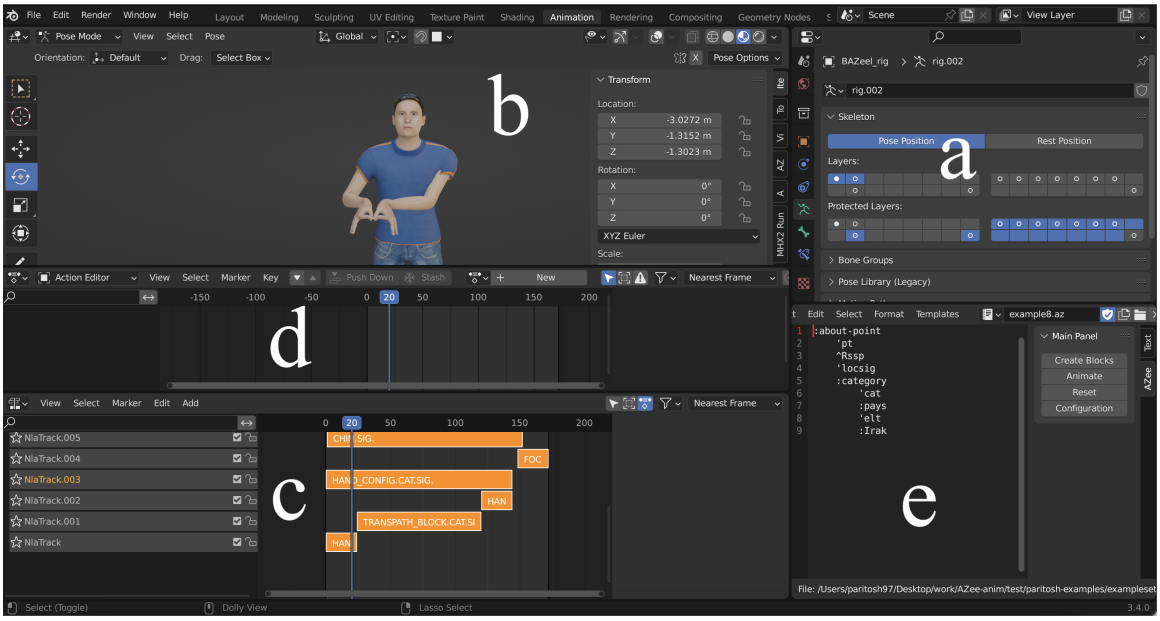

Figure 8: Blender interface. (a) Properties. (b) 3D Viewport. (c) Non-linear Editor. (d) Action Editor. (e) AZee editor

Filhol, M. and Tannier, X. (2014). Construction of a French-LSF corpus. In *Workshop on Building and Using Comparable Corpora*, Reykjavík, Iceland.

GIBET, S., LEBOURQUE, T., and MARTEAU, P.-F. (2001). High-level specification and animation of communicative gestures. *Journal of Visual Languages & Computing*, 12(6):657–687.

Holden, D., Habibie, I., Kusajima, I., and Komura, T. (2017). Fast neural style transfer for motion data. *IEEE Computer Graphics and Applications*, 37(4):42–49.

LIMSI and LISN (2022). 40 brèves. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.

McDonald, J., Wolfe, R., Schnepp, J., Hochgesang, J., Jamrozik, D. G., Stumbo, M., Berke, L., Bialek, M., and Thomas, F. (2016). An automated technique for real-time production of lifelike animations of american sign language. *Universal Access in the Information Society*, 15(4):551–566.

Mohr, A., Tokheim, L., and Gleicher, M. (2003). Direct manipulation of interactive character skins. In *Proceedings of the 2003 symposium on Interactive 3D graphics*, pages 27–30.

Pezeshkpour, F., Marshall, I., Elliott, R., and Bangham, J. (1999). Development of a legible deaf-signing virtual human. In *Proceedings IEEE International Conference on Multimedia Computing and Systems*, volume 1, pages 333–338 vol.1.

Sharma, P. and Filhol, M. (2022). Multi-Track Bottom-Up Synthesis from Non-Flattened AZee Scores. In *7th Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual & the Textual Challenges and Perspectives (SLTAT 7)*, Marseille, France.